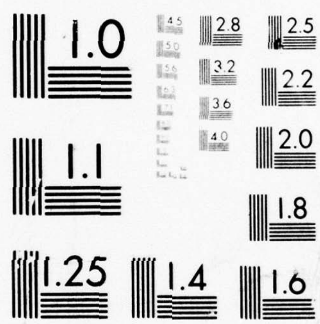
AD-A048 689 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 12/1
CERTAIN RELATIONSHIPS BETWEEN THE DISTRIBUTION MOMENTS OF BOUND--ETC(U)
AUG 77 G I YEGUDIN

UNCLASSIFIED FTD-ID(RS)T-1260-77 NL

| OF | AD A048689

END
DATE
FILMED
2-78
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

FTD-ID(RS)T-1260-77

ADA046689

FOREIGN TECHNOLOGY DIVISION

AIR FORCE SYSTEMS COMMAND

CERTAIN RELATIONSHIPS BETWEEN THE DISTRIBUTION MOMENTS OF BOUNDARY VALUES IN RANDOM SAMPLES

by

G. I. Yegudin

FOREIGN TECHNOLOGY DIVISION

EDITED TRANSLATION

FTD-ID(RS)T-1260-77 8 August 1977

MICROFICHE NR: FTD-77-C-001013

CERTAIN RELATIONSHIPS BETWEEN THE DISTRIBUTION MOMENTS OF BOUNDARY VALUES IN RANDOM SAMPLES

By: G. I. Yegudin

English pages: 10

Source: Doklady Akademii Nauk SSSR, Izd-vo Akademii Nauk SSSR, Vol 58, No. 8, 1947, PP. 1581-1584

Country of origin: USSR
Translated by: Carol S. Nack
Requester: AFFDL/FBRD

U. S. BOARD ON GEOGRAPHIC NAMES TRANSLITERATION SYSTEM

Block	Italic	Transliteration	Block	Italic	Transliteration
А а	*А а*	A, a	Р р	*Р р*	R, r
Б б	*Б б*	B, b	С с	*С с*	S, s
В в	*В в*	V, v	Т т	*Т т*	T, t
Г г	*Г г*	G, g	У у	*У у*	U, u
Д д	*Д д*	D, d	Ф ф	*Ф ф*	F, f
Е е	*Е е*	Ye, ye; E, e*	Х х	*Х х*	Kh, kh
Ж ж	*Ж ж*	Zh, zh	Ц ц	*Ц ц*	Ts, ts
З з	*З з*	Z, z	Ч ч	*Ч ч*	Ch, ch
И и	*И и*	I, i	Ш ш	*Ш ш*	Sh, sh
Й й	*Й й*	Y, y	Щ щ	*Щ щ*	Shch, shch
К к	*К к*	K, k	Ъ ъ	*Ъ ъ*	"
Л л	*Л л*	L, l	Ы ы	*Ы ы*	Y, y
М м	*М м*	M, m	Ь ь	*Ь ь*	'
Н н	*Н н*	N, n	Э э	*Э э*	E, e
О о	*О о*	O, o	Ю ю	*Ю ю*	Yu, yu
П п	*П п*	P, p	Я я	*Я я*	Ya, ya

*ye initially, after vowels, and after ъ, ь; e elsewhere. When written as ë in Russian, transliterate as yë or ë. The use of diacritical marks is preferred, but such marks may be omitted when expediency dictates.

GREEK ALPHABET

Alpha	Α	α		Nu	Ν	ν	
Beta	Β	β		Xi	Ξ	ξ	
Gamma	Γ	γ		Omicron	Ο	ο	
Delta	Δ	δ		Pi	Π	π	
Epsilon	Ε	ε	ϵ	Rho	Ρ	ρ	ϱ
Zeta	Ζ	ζ		Sigma	Σ	σ	ς
Eta	Η	η		Tau	Τ	τ	
Theta	Θ	θ	ϑ	Upsilon	Υ	υ	
Iota	Ι	ι		Phi	Φ	φ	ϕ
Kappa	Κ	ϰ	κ	Chi	Χ	χ	
Lambda	Λ	λ		Psi	Ψ	ψ	
Mu	Μ	μ		Omega	Ω	ω	

RUSSIAN AND ENGLISH TRIGONOMETRIC FUNCTIONS

Russian	English
sin	sin
cos	cos
tg	tan
ctg	cot
sec	sec
cosec	csc
sh	sinh
ch	cosh
th	tanh
cth	coth
sch	sech
csch	csch
arc sin	$\sin^{-1}$
arc cos	$\cos^{-1}$
arc tg	$\tan^{-1}$
arc ctg	$\cot^{-1}$
arc sec	$\sec^{-1}$
arc cosec	$\csc^{-1}$
arc sh	$\sinh^{-1}$
arc ch	$\cosh^{-1}$
arc th	$\tanh^{-1}$
arc cth	$\coth^{-1}$
arc sch	sech^{-1}
arc csch	csch^{-1}
rot	curl
lg	log

GRAPHICS DISCLAIMER

All figures, graphics, tables, equations, etc. merged into this translation were extracted from the best quality copy available.

CERTAIN RELATIONSHIPS BETWEEN THE DISTRIBUTION MOMENTS OF BOUNDARY VALUES IN RANDOM SAMPLES

G. I. Yegudin

(Presented by Acad. A. N. Kolmogorov on 29 July 1947)

Many studies on the laws of the distribution of boundary values and spread (order) in samples with a given volume, mainly from a normal set [1-6], have appeared recently. Furthermore, the calculations of certain constants which these reports contain are extremely significant for statistical control[1].

Footnote: [1]For example, they are necessary in the "spread" method of statistical control recommended in [7]. End footnote

If we establish the laws of the distribution of all the intermediate values, as well as the boundary values, and take them into consideration, we can find recurrent formulae which are related to each other by constants, besides certain general relationships. As we stated above, these formulae make it much easier to calculate these values. These formulae can also be considered as the relationships between certain types of integrals.

Let there be a random repeated (and in the case of an infinite general set, also unrepeated) sample of volume n from a general set whose integral distribution law is F(x). We will designate $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ as random variables which are in increasing (not decreasing) order of the values of X which form each of the samples with volume n taken: $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$.

Then it is easy to introduce $F_{k,n}(x) = \text{prob}\,\{X_{k,n} < x\}$ - the integral law of the distribution of the values of $X_{k,n}$. Actually,

the value of $X_{k,n}$ will be smaller than x in one of n - k + 1 in the following incompatible cases: 1) when all n values of X in the sample are smaller than x; 2) when n - 1 values of X are smaller than x and one value is larger than x; ...; n - k + 1; 3) when k values of X are smaller than x, and the remaining n - k values are greater than x. The probability that exactly i of the determined (let us say, first) values of X will be smaller than x and the remaining n - i values will be greater than x in a sample with volume n is equal to [prob $\{X<x\}]^i\times$ [prob $\{X>x\}]^{n-i}=[F(x)]^i[1-F(x)]^{n-i}$.

The probability that any i values of X will be smaller than x and the remaining n - i values will be greater than x in this same sample is obviously equal to $C_n^i[F(x)]^i[1-F(x)]^{n-i}$.

Thus, the unknown probability $F_{k,n}(x)$ will be:

$$F_{k,n}(x)=\sum_{i=k}^{n} C_n^i[F(x)]^i[1-F(x)]^{n-i}. \qquad (1)$$

Further, let the integrals

$$\overline{X}^r_{k,n}=\int_{-\infty}^{\infty} x^r\,dF_{k,n}(x)$$

and

$$\overline{X}_s^r = \int_{-\infty}^{\infty} x^r d[F(x)]^s = \int_{-\infty}^{\infty} x^r dF_{s,s}(x) \quad (s=\overline{1,n}),$$

exist for certain whole positive values of r, i.e., $\overline{X}_s^r$ is the initial moment on the order of r of the largest value in a sample with volume s. Then (1) results in

$$X_{k,n}^r = \sum_{i=k}^{n} C_n^i \sum_{l=0}^{n-i} C_{n-i}^l (-1)^l \overline{X}_{i+l}^r. \tag{2}$$

Now we will prove the following proposition:

The arithmetic mean of the initial moments of order r of values $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ is equal to the initial moment of the same order of the general set. i.e.,

$$\frac{1}{n} \sum_{k=1}^{n} \overline{X}_{k,n}^r = \frac{1}{n} \sum_{k=1}^{n} \int_{-\infty}^{\infty} x^r dF_{k,n}(x) = \int_{-\infty}^{\infty} x^r dF(x). \tag{3}$$

Obviously, it suffices to show that

$$\frac{1}{n} \sum_{k=1}^{n} dF_{k,n}(x) = dF(x). \tag{4}$$

in order to prove (3).

From (1)

$$\sum_{k=1}^{n} dF_{k,n}(x) = \Big\{ \sum_{k=1}^{n} \sum_{i=k}^{n} [C_n^i\, i\, [F(x)]^{i-1} [1-F(x)]^{n-i} - C_n^i (n-i) [F(x)]^i [1-F(x)]^{n-i-1}] \Big\}\, dF(x).$$

The internal sum can be transformed

$$C_n^k k [F(x)]^{k-1} [1-F(x)] + \sum_{i=k+1}^{n} C_n^i\, i\, [F(x)]^{i-1} [1-F(x)]^{n-i} - \sum_{i=k}^{n} C_n^i (n-i) [F(x)]^i [1-F(x)]^{n-i-1} = C_n^k k [F(x)]^{k-1} [1-F(x)]^{n-k}.$$

This means that

$$\sum_{k=1}^{n} dF_{k,n}(x) = \sum_{k=1}^{n} C_n^k k [F(x)]^{k-1} [1-F(x)]^{n-k} = \sum_{k=1}^{n-1} C_n^k k [F(x)]^{k-1} [1-F(x)]^{n-k} + n [F(x)]^{n-1} =$$

$$= n \sum_{k=1}^{n-1} C_{n-1}^{k-1} [F(x)]^{k-1} [1-F(x)]^{n-k} + n [F(x)]^{n-1} =$$
$$= n [F(x)]^{x-1} + n [F(x) + [1-F(x)]]^{n-1} - n [F(x)]^{n-1} = n,$$

which also proves (4), and, consequently, (3).

We will point out that equation (3) is, of course, the extension of the known property of the associativity of the arithmetic mean of n of the values to the case in question.

Using (2), we can write the following instead of (3)

$$\sum_{k=1}^{n} \sum_{l=k}^{n} C_n^l \sum_{l=0}^{n-l} C_{n-l}^l (-1)^l \overline{X}_{l+l}^r = nm_r(X), \tag{5}$$

where the initial moment of order r of the general set is designated by $m_r(X)$

In particular, if the law of the distribution of the general set is normal, then, assuming that the center of the distribution is equal to zero:

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-x^2/2\sigma^2} dx. \tag{6}$$

Then

$$\overline{X}_s^r = \frac{s}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^r e^{-x^2/2\sigma^2} \left[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-t^2/2\sigma^2} dt \right]^{s-1} dx = a_{s,r}\sigma^r, \tag{7}$$

where $a_{s,r}$, which depend on s and r, are those values whose magnitudes are necessary in statistical applications for different n, at least for r = 1, 2, 3, 4.

For the general set (6), r $m_r(X)=0$ and $\overline{X}^r_{\frac{n+1}{2}+h} = \overline{X}^r_{\frac{n+1}{2}-h}$ at all odd r. Therefore, (5) results in the following recurrent relationship

between the values of $a_{n,r}$ at different n and fixed r

$$\sum_{i=\frac{n+1}{2}}^{n} C_n^i \sum_{l=0}^{n-i} C_{n-i}^l (-1)^l a_{i+l,\,r} = 0. \tag{8}$$

Obviously, this equation can also be considered to be the recurrent relationship between integrals of type (7) at odd r[1].

[1]When examining integrals (7) at r = 1, V. I. Romanovskiy [8] established the relationship in (8) between them. End footnote

At even r, $\overline{X^r}_{\frac{n}{2}-h} = \overline{X^r}_{\frac{n}{2}+h}$ and (5) results in the following recurrent relationship between the values of $a_{n,r}$ at different h and any fixed r:

$$\sum_{k=\frac{n}{2}+1}^{n} \sum_{i=k}^{n} C_n^i \sum_{l=0}^{n-i} C_{n-i}^l (-1)^l a_{i+l,\,r} = \frac{n}{2}. \tag{9}$$

Obviously, this equation establishes the relationship between type (7) integrals at even r.

If we designate $\sigma^2(X_n) = \overline{X_n^2} - (\overline{X}_n)^2 = (a_{n,2} - a_{n,1}^2)\sigma^2 = \beta_n^2 \sigma^2$, then

$$a_{n,2} = \beta_n^2 + a_{n,1}^2,$$

and (9), not (8), results in recurrent relationships between the

scatterings of the distribution of the largest (smallest) values in samples of different volumes from the normal set.

The statistical value of the constant $a_{n,r}$ is obvious. Suppose that m random samples have been taken, each with volume n, from a normal general set. We will use $\dot{\bar{X}}_n$ and $\dot{\bar{X}}_n^2$ to designate the statistics which are the first and second sampled initial moments, respectively, with the largest values, i.e., the arithmetic means of the first and second orders, respectively, of the largest values observed in each of m samples taken, with the same volume n. Then, if we know that the center of the general distribution is equal to zero, the correlated estimate σ^2 of the scattering of the general set will be equal to the following value, as follows from (7)

$$\frac{1}{a_{n,r}} \dot{\bar{X}}_n^2. \qquad (10)$$

If the normal general set has an unknown distribution center equal to $\bar{X}$, the correlated estimate of this parameter will obviously be the value $\frac{1}{2}(\dot{\bar{X}}_n + \dot{\bar{X}}_{1,n})$ [1].

Footnote: [1]The calculation of the standard error in the approximation which gives us this estimate shows that when the researcher is limited not so much by the volume of the sample as by the difficulty of making measurements, for example, (which often occurs in

industrial applications) when determining the overall mean of the sample, it is possible to considerably decrease the number of these measurements without decreasing the precision of the approximation. Instead of measuring all N units of the sample, which is required by the usual method of measuring the mean, N - the volume of the entire sample - should be increased somewhat, then it should be subdivided into m groups of volume n, and 2m measurements of the boundary values in each group should be made in all. The boundary samples can often be selected without measuring them. End footnote

Here $\dot{X}_{1,n}$ is the smallest mean sampled value, i.e., the arithmetic mean m observed in each of the samples made of the smallest values. In this case, (10) cannot be used to estimate the scattering of the general set σ^2. Besides the obvous estimation of σ^2 by the spread of the sample, we can now also estimate the parameters of the general set σ^2 and $\overline{X}$ without considering the smallest observed values in the samples taken, basing our calculations only on the largest sampled values of $\dot{X}_n$. In this case, the value $\frac{1}{\beta_n^2}\frac{m}{m-1}\left[\dot{\overline{X}}_n^2-(\dot{\overline{X}}_n)^2\right]$, whose mathematical expectation is again equal to σ^2, should be used for the estimate of σ^2.

Furthermore, if $\overline{X}$ is the center of the distribution of the general set, $\overline{X}_n=a_{n,1}\sigma+\overline{X}$. Therefore, the following can be taken

as the estimate (not correlated) of the value of $\bar{X}$

$$\dot{X}_n - \frac{a_{n,1}}{\beta_n} \sqrt{\frac{m}{m-1}\left[\dot{X^2} - (\dot{X}_n)^2\right]}.$$

The information stated in the footnote is even more applicable to this estimate.

Received 29 July 1947

Bibliography

[1] L. H. Tippet, Biometrika, 17, 364 (1925). [2] E. S. Pearson, Biometrika, 18, 173 (1926). [3] E. S. Pearson, Biometrika, 24, 404 (1932). [4] A. T. McKay and E. S. Pearson, Biometrika, 25, 415 (1933). [5] D. Neuman, Biometrika, 31, 10 (1939). [6] E. S. Pearson, Biometrika, 32, 100 (1942). [7] B. P. Pudding and W. J. Jennet, Quality Control Charts B. S. 600 R, 1942. [8] V. Romanovsky, Biometrika, 25, 195 (1933).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER FTD-ID(RS)T-1260-77	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CERTAIN RELATIONSHIPS BETWEEN THE DISTRIBUTION MOMENTS OF BOUNDARY VALUES IN RANDOM SAMPLES		5. TYPE OF REPORT & PERIOD COVERED Translation
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G. I. Yegudin		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Foreign Technology Division Air Force Systems Command U. S. Air Force		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 1947
		13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

ACCESSION for

NTIS	White Section	☑
DDC	Buff Section	☐
UNANNOUNCED		☐
JUSTIFICATION		

BY

DISTRIBUTION/AVAILABILITY CODES

Dist.	AVAIL. and/or	SPECIAL
A		

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

12

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)